RL-TR-95-148
Final Technical Report
August 1995

BUILDING A HIGH-SPEED NETWORKING INFRASTRUCTURE

Cornell University

Malvin H. Kalos and Jeffrey Silber

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

Rome Laboratory
Air Force Materiel Command
Griffiss Air Force Base, New York

This report has been reviewed by the Rome Laboratory Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RL-TR-95-148 has been reviewed and is approved for publication.

APPROVED: Richard C. Butler II

RICHARD C. BUTLER II
Project Engineer

FOR THE COMMANDER: John A. Graniero

JOHN A. GRANIERO
Chief Scientist
Command, Control & Communications Directorate

If your address has changed or if you wish to be removed from the Rome Laboratory mailing list, or if the addressee is no longer employed by your organization, please notify RL (C3BC) Griffiss AFB NY 13441. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 1995	Final Sep 93 - Sep 94

4. TITLE AND SUBTITLE

BUILDING A HIGH-SPEED NETWORKING INFRASTRUCTURE

5. FUNDING NUMBERS

C - F30602-93-C-0214
PE - 62702F
PR - 4519
TA - 22
WU - PI

6. AUTHOR(S)

Marvin H. Kalos and Jeffrey Silber

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Cornell University
Cornell Theory Center
Ithaca NY 14853-2801

8. PERFORMING ORGANIZATION REPORT NUMBER

N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Rome Laboratory (C3BC)
525 Brooks Rd
Griffiss AFB NY 13441-4505

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

RL-TR-95-148

11. SUPPLEMENTARY NOTES

Rome Laboratory Project Engineer: Richard C. Butler II/C3BC/(315) 330-7751

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This work demonstrates the new possibilities in high performance computing, enabled by recent advances in high speed networks. The network used was NYNET, central New York's high speed network.

14. SUBJECT TERMS

NYNET, High performance networks, Supercomputing

15. NUMBER OF PAGES

40

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Project Background

In April 1993, in response to Rome Laboratory BAA 90-04, Cornell University submitted the proposal "Building a High-Speed Networking Infrastructure: An Integrated Approach to Network Management, Analysis and Prototype Applications on the NYNET Upstate Corridor." The Collaboration involved several units within Cornell, including the Theory Center, Network Resources, and the Department of Computer Science.

Cornell Theory Center is an interdisciplinary research center devoted to the study and facilitation of computational science. Under the direction of Malvin H. Kalos, the Center operates one of the fastest supercomputers in the world, and is involved in several high-speed networking projects. The Theory Center receives support from, among other sources, the National Science Foundation, New York State, the Advanced Research Projects Agency, the National Institutes of Health, and the Center's Corporate Research Institute.

Network Resources is a unit of Cornell Information Technologies. CIT has provided world-class leadership both in deploying advanced networking technologies on campus and in designing and influencing the design of leading networking protocols and algorithms. As part of its role in the national computing arena, CIT is a partner with the Cornell Theory Center in high technology projects, including NYNET.

CIT currently monitors and operates international connections for NSFnet, the local T3 NSFnet backbone and regional connections, the campus FDDI backbones, 300 routers, and connections for more than 10,000 campus data users. CIT is also responsible for a 16,000 line PBX. The department has extensive development efforts in distributed network applications, desktop video, and network routing, and is involved with various national task forces, standards groups, and industry partners.

The Department of Computer Science in the College of Engineering at Cornell will be using some of the local NYNET resources developed, in part, for this project. Among the uses will be research and development for ISIS and HORUS, under the direction of Professors Ken Birman and Thorsten von Eicken.

NYNET is an ATM-based network testbed involving, among others, NYNEX, Rome Labs, Cornell, Syracuse, Columbia and Polytechnic Universities. The original white paper describes NYNET as follows:

> "NYNET is a high-speed fiber-optic communications network linking multiple computing, communications, and research facilities in New York State. NYNET incorporates what was formerly known as the Supercomputer Corridor, combining it with the developing Downstate Corridor. This broader emphasis underscores NYNET's statewide reach and broader applications base, initiating a vital communications and computing infrastructure for the state of New York. Whereas a traditionally voice-based communications infrastructure has traditionally served as the primary environment for collaborative efforts, NYNET will become a state-wide high-bandwidth channel of all multimedia communications (voice, video, and data) and supercomputing capability."

Since its original inception NYNET has been expanded to include other partners in downstate New York, and Cambridge, Massachusetts.

This project is the first step in an expanded Cornell–Rome Labs relationship. This project funds one portion of a far broader scope of activities. The work described below encompasses numerous projects, only a portion of whose support came from this contract. The preliminary results of this effort were presented in a project briefing and demonstration at Rome Laboratories on September 21, 1994. An opportunity was provided for input from that briefing to be incorporated into this final report. The headings in the report are keyed to the Tasks/Technical Requirements outlined in the Statement of Work.

4.1.1.1 Interaction between ATM and other protocols

The effect of lost ATM cells on higher level protocols is not a new problem, but one that occurs in any case where larger datagrams are fragmented for transmission and one or more fragments is lost. The small cell size of ATM coupled with high bandwidth links greatly exacerbates the problem. Factors that could cause cell loss are: link errors or cells "dropped" due to congestion. Our main interest is in congestion because, with appropriate management tools and techniques, the impact can be minimized. The primary issue here is that, if one or more cells from a packet are lost, the rest of the cells become useless data, although they still contribute to the aggregate congestion.

The nature of the NYNET Trial has not been particularly conducive to primary research on these issues. First, because NYNET is intended to be an applications development environment using production, vendor-supported ATM hardware and software, our ability to make changes to the ATM fabric has been minimal. Second, the six-month delay in connecting Cornell allowed others time to study this in considerable depth. We have followed this literature closely and summarize it here. Although there is no single, standardized solution, we feel the problems, approaches, and solutions are now fairly well understood.

A landmark work on this topic is Allyn Romanow and Sally Floyd's paper: "Dynamics of TCP Traffic over ATM Networks." (available by anonymous ftp from playground.sun.com: pub/tcp_atm/tcpatm_extended.*.ps). Although their study focused on TCP in a local (low latency) environment, most of the work is directly applicable to other protocols and is extensible to wide-area (high latency) networks.

The goal is to cost effectively transport connectionless packets over ATM networks. In a degenerate case, one could simply allocate Constant Bit Rate (CBR) connections of sufficient mesh and bandwidth to guarantee that there would be minimal congestion. This would clearly be extremely costly and cumbersome. The goal, therefore, is to use Available Bit Rate (ABR) services to best advantage. An option is to allocate large enough output buffers such that one can buffer the rest of the data that is sent during the time it takes for the sender to recognize it has not gotten an ack and stop sending. This means a buffer greater than twice the network bandwidth-latency product. The buffer approach has deficiencies in terms of cost (both hardware and additional latency) and the effectiveness is offset as one increases the number of contending connections.

Floyd and Romanow tested, via simulation, two alternatives: Partial Packet Discard (PPD) and Early Packet Discard (EPD). With PPD, if a switch must drop a cell, it drops the rest of the cells associated with the same VC until it sees a cell with the AUU parameter in the header set to indicate the end of the AAL packet. Since AAL5 does not allow multiplexing of packets on a single VC, this can be used to identify the last cell from that packet. The switch must maintain state tables on a per-VC basis for which are using AAL5 and want to use PPD. It must also keep the state for which VCs are

currently having their cells dropped. A weakness of this approach is that cells are only dropped after the buffer has begun to overflow, and when a cell has been dropped previous cells associated with this packet may have already been queued or sent. In their implementation of EPD, the switch drops all cells associated with a particular VC whenever the portion of the buffer in use exceeds a pre-set threshold. Since EPD does not require cooperation between ATM switches or end-to-end congestion control, it applies very well to protocols other than TCP.

In conclusion, we feel that we cannot make changes to existing, legacy protocols to accommodate ATM. ATM switches and protocols must allow for efficient and transparent transmission of existing protocols without introducing excessive packet loss. Further study is needed to determine the most efficient physical buffer sizes, proportion of input to output buffering, and thresholds for EPD as well as interactions between EPD and other ATM control mechanisms such as Random Early Detection (RED).

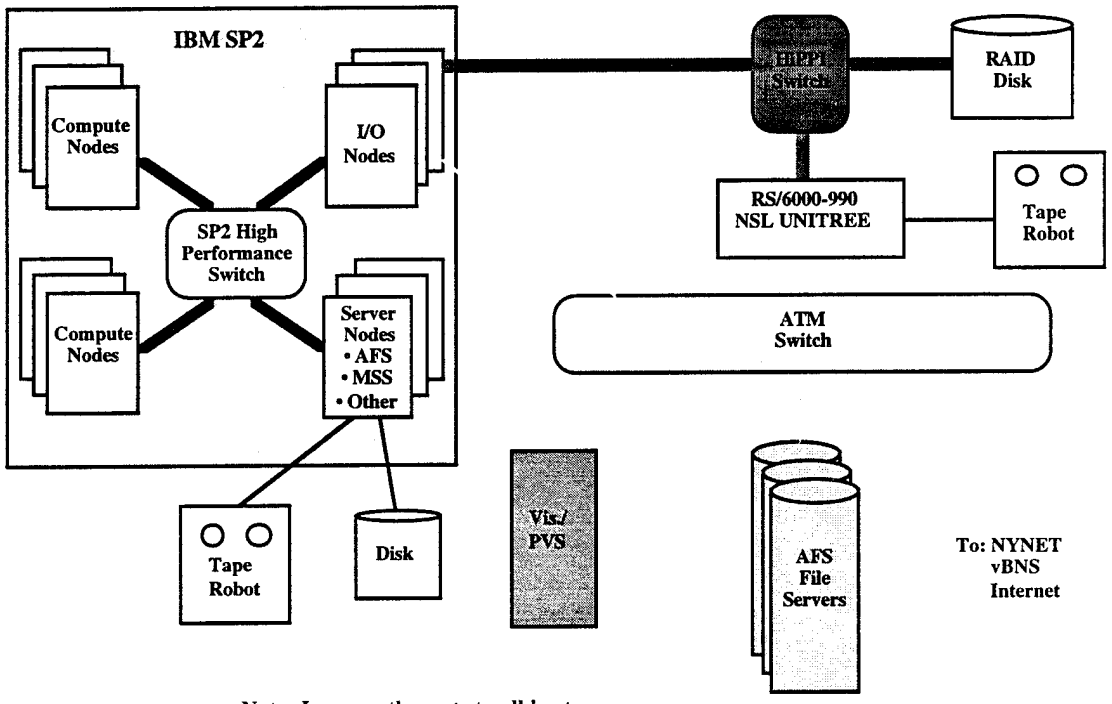
4.1.1.2 Interoperability issues

There are numerous interoperability issues between different ATM implementations. The lack of a standard for SVCs has been a major inconvenience. Use of PVCs for every connection between any two machines has required excessive personnel time and coordination, since different groups have responsibility for various workstations and switches. We have, therefore, relied heavily on Fore's proprietary SVC implementation: SPANS. SPANS has caused us interoperability problems when going through other vendors switches since it requires specific VPI/VCI combinations which may not be available. Specific incidences are: when we upgraded our ASX-100 software to version 2.3.2, we discovered that they had changed the dynamic allocation of VCs used by SPANS from sequential to random in the range of 32 to 255. This required allocation of all of these VCs on the NYNEX switch in Syracuse to enable SPANS to work and made it impossible to connect downstate via SPANS through Wiltel's NEC switch since some VCs in that range were reserved for their own uses.

4.1.1.3 Interconnect ATM with other LAN Technologies

The Theory Center has introduced ATM into its local network environment, which includes ATM, HiPPI, FDDI, and ethernet. A copy of the Center's network-based environment diagram appears below. We expect to continue enhancing this configuration, particularly in the ATM arena. The Center, which already runs a Fore ASX-100 ATM switch (soon to be upgraded to an ASX-200), will be acquiring a 50-port IBM 8260 ATM switch. In addition, as part of the Center's NSF-funded vBNS effort, the Center expects to have a NetStar ATM-HiPPI GigaRouter.

CORNELL THEORY CENTER
Configuration (1H95)



jas
11/13/94

4.1.1.4 Explore the viability of ATM as a campus LAN technology

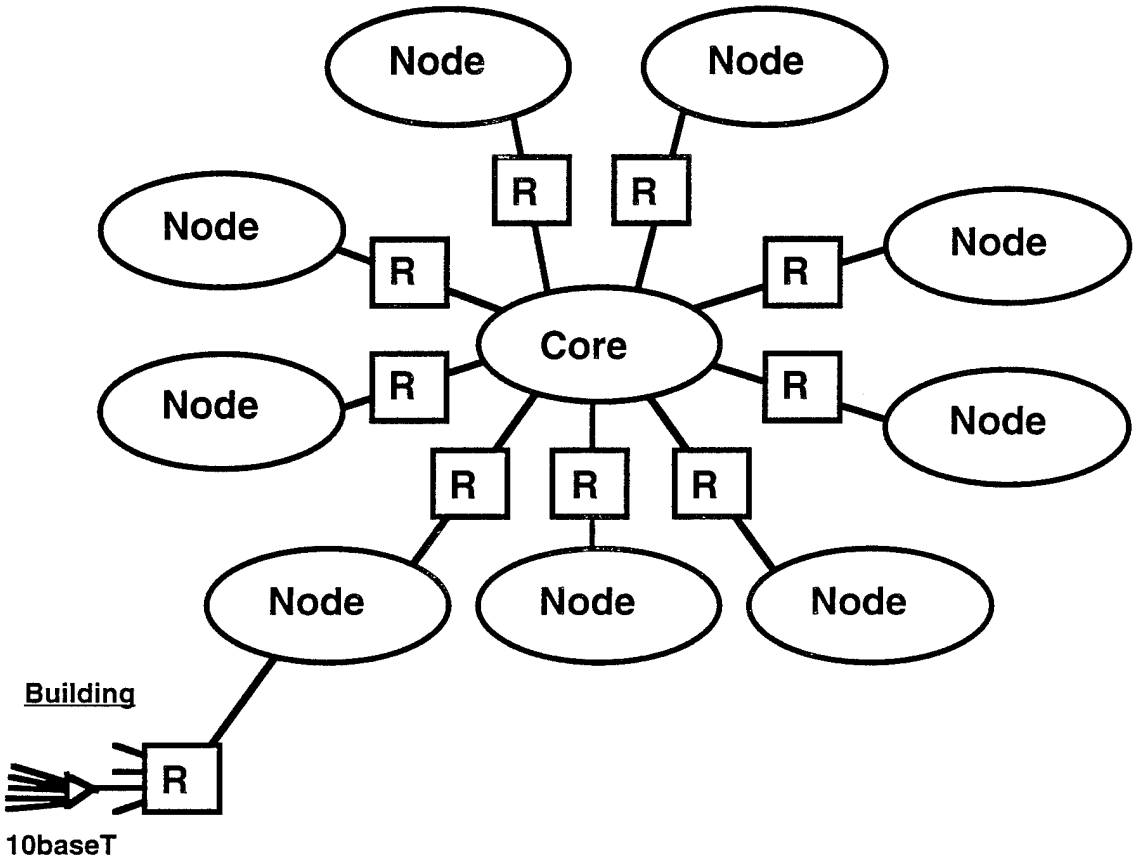
Cornell has been exploring the viability of ATM technology to create a campus communications infrastructure that could provide data, voice, and video transmission. Cornell's current funding models for communications infrastructure will not support construction and maintenance of three separate technologies and systems for providing these services. ATM shows the greatest promise for technology that may allow integration of these services, but the maturity of the technology—specifically in areas of capabilities, usability, manageability, and affordability—must be considered. Also, a migration plan must take into account the installed base of existing equipment which has not yet been fully amortized.

To fully explain our migration plans, we begin by describing our current communications infrastructure and installed base. Cornell's telephony infrastructure consists of an AT&T Definity PBX with approximately 15,000 lines. The Definity is distributed with a central module and 8 remote modules serving sections of the Campus and connecting to the central module via fiber optic links. The topology of the data network backbone parallels this, since it uses fiber optic links bundled with the telephone fiber. There is no comparably ubiquitous system for video distribution. There is a 300 MHz mid-split CATV system which was installed for data communications and which touches 30 buildings and carries five channels of TV originated from our TV studio or received by satellite dish. We have also used multi-mode fiber for point-to-point, single channel video links for live projection of events to overflow areas. We have also been experimenting with desktop video conferencing over the Internet (see other sections on

CU-SeeMe). We will focus here on the current data infrastructure, since it will be the primary building block toward an integrated system.

Cornell currently supports about 9,000 LAN ports on about 500 LANs, routed to the backbone. Most of the LAN connections are via 10baseT ethernet utilizing additional pairs of Category 3 UTP installed with the telephone wire. A typical LAN is a collection of jacks within an administrative workgroup, sharing a 10baseT hub or concentrator at the Building Distribution Frame (BDF). This LAN is then connected to a local router in the building on a "Node" FDDI network which is routed to a "Core" FDDI Network as shown below.

Cornell FDDI Backbone - Current



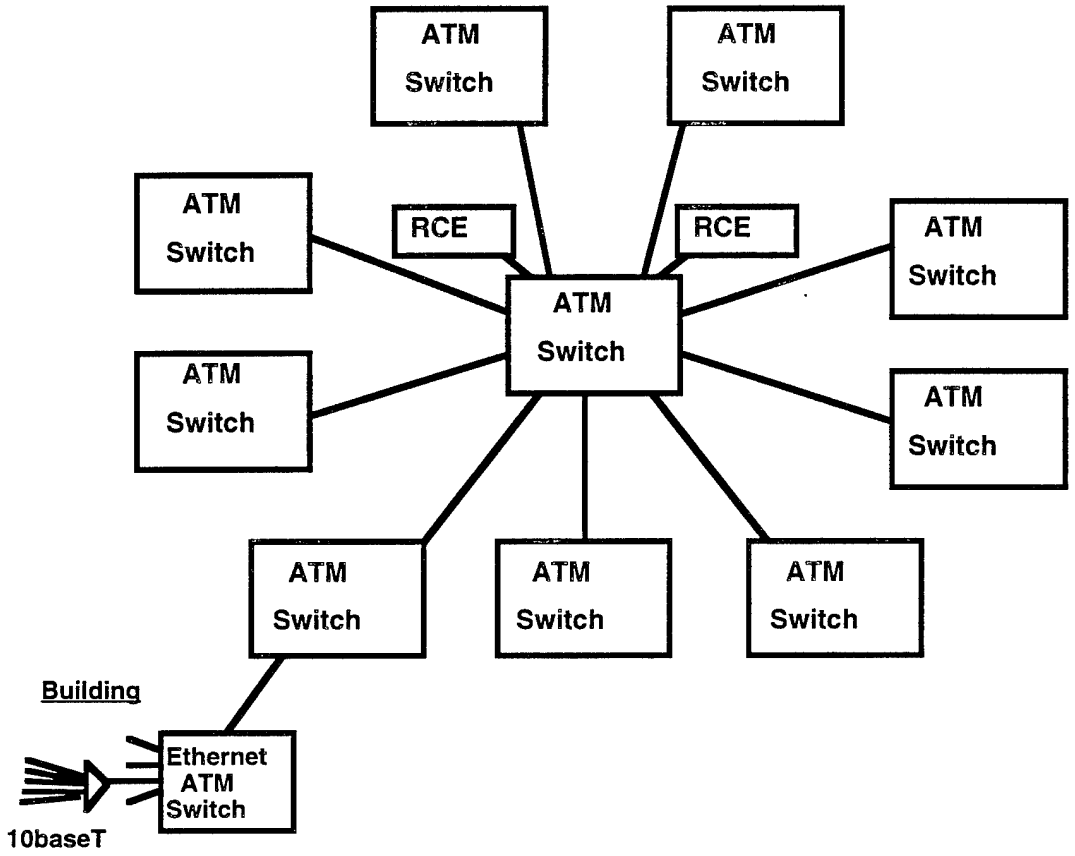
There are about 110 buildings connected to these FDDI networks. The routers are IBM model 6611+ and are a joint development of Cornell and IBM. The FDDI connections are DAS and are star wired to dual-homed concentrators for redundancy and reliability. This architecture allows us to decrease the number of nodes on a shared ethernet for increased performance or to add extra FDDI interfaces into building routers to provide user FDDI LANs (as with the Cornell Theory Center) for special high-performance needs. The only protocols transported across the backbone are TCP/IP and Appletalk. Monitoring and remote problem diagnosis is via SNMP and all devices (including wiring concentrators) must be manageable via SNMP. The backbones, building routers, and

hubs are all physically secured so that applications can be assured that a connection from a particular IP subnet does, in fact, originate from that physical LAN. Thus routing (as opposed to bridging) at the periphery to the user LAN is essential for security and desirable for manageability and control.

An obvious potential bottleneck in this architecture is the aggregate traffic of nine 100 Mb "Node" networks feeding into a single 100 Mb "Core" network. We have explored various options for higher performance in the core, including ATM technology. MAC level bridging between Node FDDI networks is acceptable from a security perspective, since no user LANs connect directly to them. Our conclusion is that, at the present time and probably for the next two years, the most efficient and cost effective solution to replace the Core FDDI is a DEC Gigaswitch. This would replace the dual FDDI concentrators in the Core and the FDDI to FDDI routers at the Nodes. We would extend the Node FDDI networks back to the central core and bridge these FDDIs using the Gigaswitch. The ATM alternative would require a 9-port switch at the core and nine ATM to FDDI edge routers at the nodes. We do not believe this can be done as cost effectively or reliably as with a Gigaswitch collapsed backbone and there are no advantages in terms of additional functionality from ATM with just tying together existing packet-based LANs.

We propose Cornell's migration to ATM technology be done in three phases spanning five to six years: Phase 1: replace the FDDI backbones, Phase 2: take ATM to the desktop, and Phase 3: integrate voice in the ATM network. These phases are not sequential, but will be implemented as funding and maturity of the technologies permit. Phases 2 and 3 will likely be prototyped long before Phase 1 is begun.

Cornell ATM Migration - Phase I

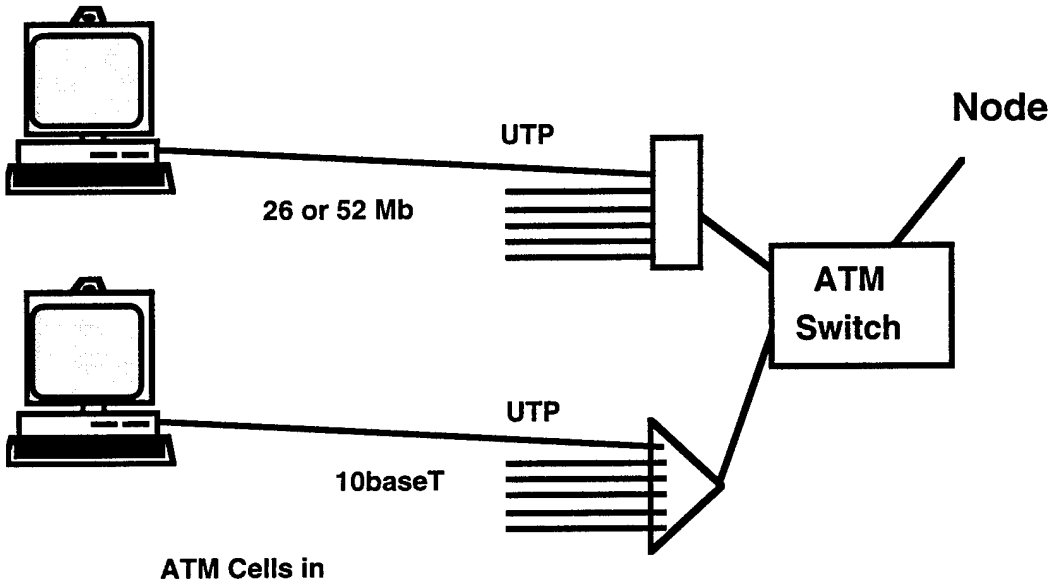


The goal of Phase 1 is to create what we term the "500 port (virtual) router." Since core links can be higher speed, there will be considerable aggregate bandwidth gain and central bottlenecks will be overcome. This will also create the base infrastructure to build Phases 2 and 3. The "RCE" in the above diagram is the "Route Computation Engine" or route server for legacy packet-switched protocols. We would certainly require TCP/IP and probably Appletalk routing. It is important to note that we will not consider a network using only "ATM LAN Emulation." It would not allow the security we require nor, we believe, be manageable. Also, we do not find the prospect of legacy (physical) routers between emulated LANs very appealing. Each ethernet port would need to be able to handle full ethernet standard MAC addressing. The RCE would need to treat these ethernet ports as separately routed networks and be able to communicate with other RCEs and routers using current routing protocols (e.g. OSPF, BGP4). This network could be built today with proprietary technology (e.g. Newbridge Networks' Vivid Yellow Ridges) with compliance to today's standards (UNI 3.0 and RFCs 1483 and 1577), but it is unlikely that it would interoperate with another vendor's interpretation of these standards. Even at today's prices, this network could be built for a small fraction of the cost of the equivalent network using routers and FDDI. However, there does not yet exist the richness and variety of tools available to manage today's packet networks. We believe it will be cost-justified and possible for us to install a manageable network of this type by late 1996.

Cornell ATM Migration - Phase II

(ATM to the Desktop)

Building Distribution:

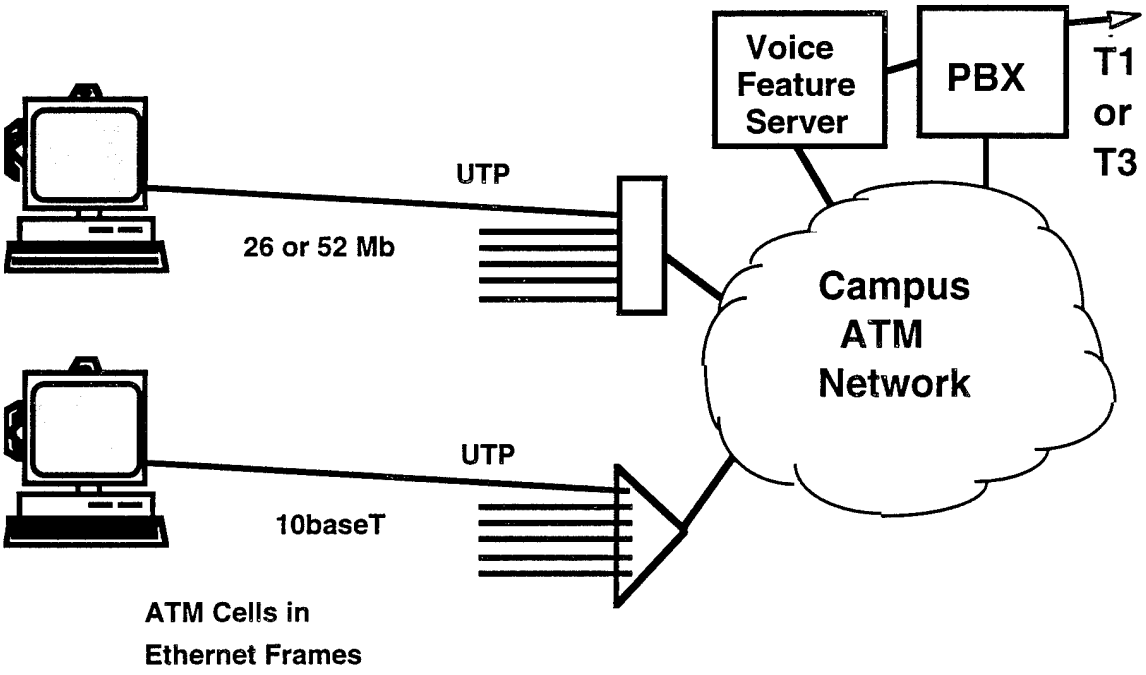


Phase 2 will bring ATM technology to the desktop. One possible option for this is IBM's 26 Mb over UTP proprietary technology. However, a major limitation to this technology for us is the 100 meter limit over Category 3 wiring. We have approximately 16,000 lines of Category 3 wire and estimate that 20% of these exceed 100 meters. Thus a ubiquitous installation of this type would require a major rewiring of the campus, at a cost beyond what we believe we could fund. Another alternative is the 16-CAP UTP standard at 52 Mb proposed by AT&T and approved by the ATM Forum. This strongly appeals to us since it will run longer distances over Category 3 wire, although at decreased speed. A third option is one we have proposed to vendors to run ATM cells in ethernet frames (see Appendix A for more detailed description). This would leave the current 10baseT NIC card in the workstation and provide a software interface to allow applications access to ATM services. By a couple of simple rules: only one node per mux port and only allowing cells for one VPI/VCI in any given ethernet frame, it is possible to eliminate the extra addressing overhead of ATM in the frames and reach usable bandwidths of 7 to 9 Mb. This is adequate for a CATV-quality MPEG-2 video stream, with plenty left over for voice and data. The mux which replaces a shared-ethernet hub, passes all cells up a higher bandwidth (e.g. 155 Mb) connection to an ATM switch, so it does not need ATM switching and can be a fairly simple (i.e. inexpensive) device. We expect to produce a working prototype using ATM cells in ethernet frames in early 1995. We believe this solution could be procured in 1996-97 for $100 per port. This could be accommodated within our existing rate structure without a cost increase to users. While we will certainly have users needing higher bandwidth (26 to 52 Mb or more), this would be offered as a premium service at a higher cost. We estimate $600 to $1,000 per connection including hub, NIC, and installation. We believe, given the cost

difference, that ATM over dedicated 10baseT links will satisfy the vast majority (est. 90%) of our users for at least five more years.

Cornell ATM Migration - Phase III

(Voice Integration)



Phase 3 of Cornell's ATM migration strategy integrates voice over the ATM network and begins eliminating the separate voice PBX infrastructure. The difficulty in using ATM for voice is not in the bandwidth requirements (the bandwidth is comparatively low--64 Kb) or in the guaranteed bit rate (Constant Bit Rate is no problem for an ATM network), but in providing the same level of features, functionality and sound quality PBX users are accustomed to today. Some examples of features used today are multiple line appearances, call forwarding, multi-party conferencing, and voice mail. Within a few years, we expect "shrink wrap" voice applications for workstations to be commonplace. The missing pieces are a centralized "Voice Feature Server" to provide PBX-like features and function and protocols to talk between the server and clients. We expect it will be most efficient to maintain a small PBX (connected to our ATM network) to connect "dumb" phones (e.g. emergency phones) and to provide efficient least-cost call routing to long-distance lines and handle billing functions. We expect to prototype voice over ATM within one year and begin phasing this technology in and our PBX out over two to five years.

4.1.1.5 Identify network bottlenecks

Network bottlenecks in ATM generally only appear in Available Bit Rate (ABR) service where there is contention for a finite resource. Since a standards-based ABR has not yet been defined and we have very few nodes on NYNET to provide aggregate loads, there has not been opportunity for identifying or even creating any real bottlenecks. Once the ATM Forum has defined ABR service and resolved flow control issues, this will require future study.

4.1.1.6 Explore the use of various routing protocols

There are two competing approaches to providing traditional packet-based data services over ATM: LAN Emulation and Classic IP (see RFC 1577). In LAN emulation, current (physical) routers can be used to connect between these virtual LANs and the routers can use traditional routing protocols between themselves and need have no awareness of the actual network technology (i.e. ATM). We believe that routing of packet data (as opposed to bridging as in LAN emulation) is necessary for a campus-size securable and manageable network. With an ATM network, the routing function can be separated from the cell forwarding function with the route calculated and served from a central server. The centralized route server or route computation engine (RCE) will need to communicate status of its virtual connections and routes to other physical or virtual routers (e.g. RCE). We see no reason why current IGPs (Interior Gateway Protocols) are not adequate to this task within an Autonomous System (AS). Obviously, more modern routing protocols with enhanced features (like OSPF over RIP) still provide better functionality in this environment. For Exterior Gateway Protocols (EGPs) between ASs, the current versions of ATM do not allow a standards-based SVC facility so we are limited to PVCs. These PVCs are treated just like real physical links and any current EGPs work perfectly well.

4.1.1.7 Develop applications

Cornell has developed and demonstrated applications that are described in sections 4.1.6.1 and 4.1.6.2. In addition, Cornell has an experimental ATM-attached WWW server (berkshire.tc.cornell.edu), and plans on adding ATM capabilities to its full WWW (www.tc.cornell.edu) server in the near future.

4.1.1.8 Effects of Latency

Network latency poses special problems for distributed applications. As network speeds increase, the bandwidth-delay product increases and the amount of data in transit increases. Flow control messages from the receiver to the sender must transit back through the network path- thus it can take a full round-trip time for a sender's response to a network congestion situation to take effect. Therefore, to insure that no data is lost, one must buffer an amount of data equal to twice the bandwidth-delay product. Traditional means of insuring database synchronicity such as "two-phase commit" for DRDBMS (Distributed Relational Data Base Management System), which are already considered costly, may prove to be completely uneconomical in a wide-area ATM network environment.

The Center is planning on using ATM networks, such as NYNET, to connect multiple SP2 supercomputers and distribute applications across them. Latency will be particularly important in such an environment, and these experiments will involve further study of this topic.

4.1.2.1 Add SNMP Agents

We have enabled the SNMP agents for monitoring of the three Fore ASX-100 ATM switches Cornell presently has in operation. We have given access to these to other NYNET participants. We currently do not have any edge routers installed.

4.1.2.2 Modify Local Management Information Bases

We have imported Fore's proprietary MIB information to a workstation running IBM's NetView/6000. This station is in the Cornell Theory Center's Operations area and is monitored 24 hours a day. In addition to monitoring Cornell's three ASX-100 switches, we are monitoring Rome and Syracuse University's ASX-100 switches and Cornell's two SGI workstations. See copies of NV/6000 network maps in Appendix C.

4.1.2.3-2.5 NYNET Exploration Topics

The Center and NR have been active with other NYNET participants and the Applications and Infrastructure committees, which are exploring these topics. In addition, Cornell has worked to enable use of its facilities to NYNET participants for such exploration. Actual solutions for these problems are broad research topics in the academic, industry and military communities.

4.1.2.6 Integration of management tools

Cornell's Network Resources division manages a network with about 200 routers, 500 LANs, and 10,000 end nodes. We have been trying to migrate and integrate our network monitoring and management onto IBM's NV/6000. However, due mostly to difficulties resulting from our non-standard routers, we have not been able to migrate successfully or reliably. We are investigating Cabletron's Spectrum as it appears to be the best integrated network management tool available. It should allow dynamic actions as a result of network alarms and statistics. Our production monitoring station still runs a SNMP map tool from PSI and NYSERNET. It is difficult to configure and non-hierarchical, but works and reports reliably. This is not used for network management, but solely monitoring. Trouble tracking and problem resolution are still a manual process. Statistics gathering and remote problem diagnosis using SNMP are done using locally written Unix scripts.

4.1.3 Multiparty conferencing

Cornell University has developed a desktop video conferencing application called CU-SeeMe. The application was originally developed for Macintosh computers and has now been developed for PCs running Microsoft Windows. The applications use TCP/IP for transport and are publicly available via anonymous ftp from gated.cornell.edu. They can be used from anywhere an Internet connection is available. The goal has been a low cost

service. To receive a conference simply requires obtaining and running the application. To send a conference requires a digitizing board called a frame grabber and a camera which together cost less than $500. Audio is handled by an application developed at NCSA called Maven and has been directly incorporated with permission and support from Charlie Kline, the author. Multi-party (more than two parties) conferencing currently requires a UNIX machine running an application we have developed called a "Reflector" which combines and retransmits the video streams to multiple recipients. The Reflector software is also available via anonymous ftp from gated.cornell.edu. The project was begun in mid 1992 and has received funding from the NSF since October, 1993.

4.1.3.1 Compression and transport

Since most of the Internet of today has relatively low bandwidth (e.g. T1 or 1.5 Mb links), extensive compression of video streams and control over the bandwidth used was required to make CU-SeeMe functional. The compression scheme is proprietary since none of the current standards-based algorithms could meet these requirements or be accommodated with the limited CPU power available. However, a CU-SeeMe decoder has been incorporated in NV (the public-domain Internet desktop video system for UNIX systems). The system first eliminates every other pixel (in Large screen mode) resulting in a 4:1 compression or a 16:1 compression in Small screen mode. Next, color information is eliminated and luminance information is reduced to 4 bits per pixel. Then, inter-frame differencing of 8 by 8 pixel squares is performed and only the changed squares are sent. Finally, a lossless compression gains an additional 50%, using a prediction/correction coding based on spatial redundancy. Judging from the thousands of people we know of that have installed this software and paid to equip their machines to send (some have even bought new Internet connections to run CU-SeeMe), even this lossy compression has been found to be very useful — i.e.. low-quality video is better than none at all. We have found that people will readily accept low quality video that is instantly available.

Bandwidth limiting is first set by the sender with a maximum and minimum cap. The receiver reports his rate of received packet loss back to the sender- if the loss exceed 5%, the rate is lowered and the rate is raised if the packet loss rate is less than 5%. Thus, the bandwidth is dynamically adjusted based on packet loss.

All CU-SeeMe streams are sent using IP. We have been using Fore Systems' IP drivers on a UNIX system ("Skyhawk": a Sun IPC with a Fore SBA-200 NIC) running the Reflector software for CU-SeeMe over NYNET.

"Closed groups" have been implemented by allowing a "conference id" to be required to join a conference. Thus, even if you have the IP address of the Reflector for a conference, you must also know this code. Users have not expressed a strong need for encryption. However, this may be required in the future — particularly for medical applications.

4.1.3.2 User interface

The user interface for CU-SeeMe has seen considerable development and improvement. The user can set switches to receive all calls or can acknowledge and allow them individually. Multiple screens of other participants are handled via the Reflector. A "Participants Menu" is provided. The user can open additional windows from this list or

can eliminate them by simply closing their window. If a user closes a window from another participant, the Reflector is informed and it stops sending that stream, thus "pruning" the stream and not wasting network bandwidth. The current practical limit for a multi-party conference is eight active participants. Under development is the option of using "thumbnail" windows (30 x 40 pixels). Participants one just wants to monitor would still be there, and this would allow many more participants.

4.1.3.3 Intelligent bridge

The Reflector is the UNIX application which allows intelligent multi-party CU-SeeMe conferences. See previous sections.

4.1.3.4 Linkages to instructional material

CU-SeeMe has incorporated multi-media using an auxiliary data feature, a reliable one-to-many transport and a plug-in architecture permitting extension of CU-SeeMe functionality with separately compiled component modules. Ultimately, these extensions will allow conference participants to distribute nearly any other materials to other participants. These materials could be anything from drawings to spreadsheets to MPEG movies. One use being made of these facilities is in telemedicine, specifically telepathology via a plug-in developed at CUMC. The alpha version of CU-SeeMe which was demonstrated at Rome also incorporates a "Slide Window." The Slide Window allows you to send a single still image of a single frame of video at full NTSC resolution. This, depending on available network bandwidth, may take some time. Once the image is received, both participants can manipulate a shared pointer. The alpha version should be fully tested and released by January 1995.

4.1.4 Digital library infrastructure

The Cornell Digital Library project is a collaborative project supported by Xerox Corporation, Sun Microsystems, the Commission for Preservation and Access, and the New York State Program for the Conservation and Preservation of Library Research Materials. The goal is to preserve deteriorating books in digital form while creating a paper copy and to allow printing on demand of the stored books. Endangered books are scanned at 600 dpi resolution and stored on either magnetic or optical media. The digital book images are managed and accessed via a custom server application running on a Sun workstation attached to Cornell's network and accessible from the Internet. The books can be printed on a high-speed Xerox Docutech printer, which is also attached to the network, at full 600 dpi resolution. Currently, around 2,000 volumes have been digitized from Cornell collections. They range from volumes in mathematics and other subjects to New York State local history. Client software was developed for Unix, DOS, and Macintosh workstations. The client was designed to view the images at a resolution of 100 dpi to accommodate today's video display technology and to provide reasonable response over today's networks. The client is intended to be used to browse the books to determine if it is really what the user wants and, if so, the book can be printed on demand. With a very high-speed network such as NYNET, it is possible to "lip through pages with the client almost as quickly as with a real book. The client and server applications are available on the Internet by anonymous ftp from library.cornell.edu.

The Theory Center maintains an extensive digital repository on its WWW server. This repository includes text, graphics, animations, and sounds. The Center strives to make all

of its user materials (documentation, information, forms, etc.) available via this mechanism. Where ever possible the Center tries to make researchers' results (animations, abstracts, technical reports) available through this server as well.

4.1.5.1 Medical applications on NYNET

Cornell University Medical College has developed extensions to CU-SeeMe which permit it to be used for remote medical teaching, research and diagnosis. (See CU-SeeMe plug-ins described in section 4.1.3.4). This system was recently demonstrated at Rome Laboratory and will not be detailed further here.

4.1.5.2 Medical imaging

Medical viewing by magnetic resonance imaging (MRI) has proven essential for diagnostic evaluation of trauma and disease. In this application demonstration we have shown how a combination of computation, computer graphics, and network access allows evaluation of medical data which extend the capabilities of MRI instruments. Test MRI data from patients at Brigham and Women's Hospital in Boston was formatted for display as 2D slices, then converted into 3D volume data. The resulting 3D data was then available for reformatting into slices at arbitrary angles and for 3D volume rendering. As the analysis proceeded, the investigator was able to rapidly and interactively modify the image viewing parameters (e.g. viewing angles, lighting), the image content (e.g. contrast, slicing, edge enhancement), and view the resulting information in stereo. The use of standard UNIX platforms (rather than specialized MRI image computers) has allowed us to easily connect to the network and modify the image content in novel ways.

These, and other medical images can be found on the Center's WWW server at: http://www.tc.cornell.edu:80/Research/Articles/BIO/MCB/BioMed/.

4.1.5.3 Visual feedback and transmission latency

We have found that the visual display of medical information from a remote site can be accomplished in a variety of ways falling between two extremes. Consider a remote medical imaging facility capable of transmitting information to a physician's office. On the one extreme, is the direct transmission of digital images. In other words, the computation and conversion of the medical data into images is performed entirely at the remote site with only the final pictures being transmitted and displayed on a "dumb" console in the physician's office. At the other extreme, raw data collected at the remote site can be sent to a "smart" console in the physician's office where computation and image rendering can be performed locally. The bulk of the processing can thus reside on either side of the network or even be shared between the two.

We have constructed a networked visualization system capable of exploring different ways of distributing the processing work between connected sites and evaluating how well these choices deal with the problem of transmission latency. Our system is a combination of a commercially supported scientific visualization program (IBM Visualization Data Explorer), an NSF MetaCenter-developed virtual reality program (the CAVE and CAVE-simulator code written at the Electronic Visualization Laboratory at the University of Illinois) and supporting code written at the Cornell Theory Center.

Stereo presentation and immersive 3D displays (virtual reality)

Since medical data sets are inherently three dimensional and physicians rely upon their understanding of the shapes and structural features in the body to make their diagnosis and plan treatment, the appropriate use of depth cues[1] in visualization has immediate advantages:

- Resolution of ambiguous features. Magnetic resonance imaging itself does not always resolve internal boundaries very well. Knee cartilage, for example, not only has a diffuse appearance on an MRI but also represents a highly nonplanar object for which individual slices and projections may offer a misleading notion of shape. Motion parallax gained with a highly responsive system or with the use of stop-frame animation helps to resolve ambiguities.

- Improved ability to locate small features in large, complex fields of view. Our experience in immersive 3D displays (the CAVE at the Electronic Visualization Laboratory) has dramatically illustrated how a well-studied molecular surface can, in fact, yield new, previously unseen features. The Acetylcholinesterase enzyme, a target of much drug-design research, has a long tunnel which leads to the active site cavern at the center of the molecule. Our model of the substrate, acetylcholine, was thought to be partially embedded in the molecular surface, with the acetyl end penetrating into the solvent-inaccessible region. When examined in the virtual reality environment of the CAVE, with the viewer standing in the cavern, a solvent-sized hole in the surface was discovered that actually exposed the acetyl end. The ability of the viewer to explore the data from a unique and normally inaccessible angle yielded a surprising new observation. The same would equally apply to the location tumors, or to the examination of blood vessels from inside the eye.

Issues in applying virtual reality to real-time remote imaging

In our experience, the major issue in the application of virtual reality to remote imaging is network speed. If interactivity is essential to a diagnosis, then it is the transmission speed of the network that determines the mode of operation. If bandwidth is high enough to sustain an image update rate of 2-3 frames per second, then direct image transmission is effective, especially if very large data sets are involved. Practically, however, network bandwidth is a limiting factor, especially if images need to be of high resolution (1280x1024 pixels of better) and full color (24 bits per pixel). We have found that if

---

[1]
- Stereopsis: each eye sees a different view of the object and the brain extracts three dimensional information.
- Atmosphere: points on the object closer to the viewer are brighter and clearer.
- motion parallax: points closer to the viewer move faster (e.g. rotation).
- Perspective: parts of the object closer to the viewer seem larger.
- Focus: only points at the focal length are in focus.

high-speed interactivity is needed, as it is to provide motion parallax depth cues, then the best approach is to transmit the data and render it locally in the physician's office.

Medical data sets can however be quite large, so the optimal solution is to preprocess the data at the collection site (the MRI machine, for example) and send a reduced-size data set or geometry based on what the physician actually needs to see. An example would be a large MRI volumetric data set. The physician may only need to see a color-coded surface of a particular organ. Or software and create this surface representation on site and transmit it to the physician's workstation for interactive manipulation. Since the data are cached on the local workstation, the physician can continue to manipulate and explore the data while waiting for the next updated information to arrive.

4.1.6 Demonstrations

The following were demonstrated at Rome Laboratories on September 21, 1994:

4.1.6.1 Virtual x-ray space and image analysis

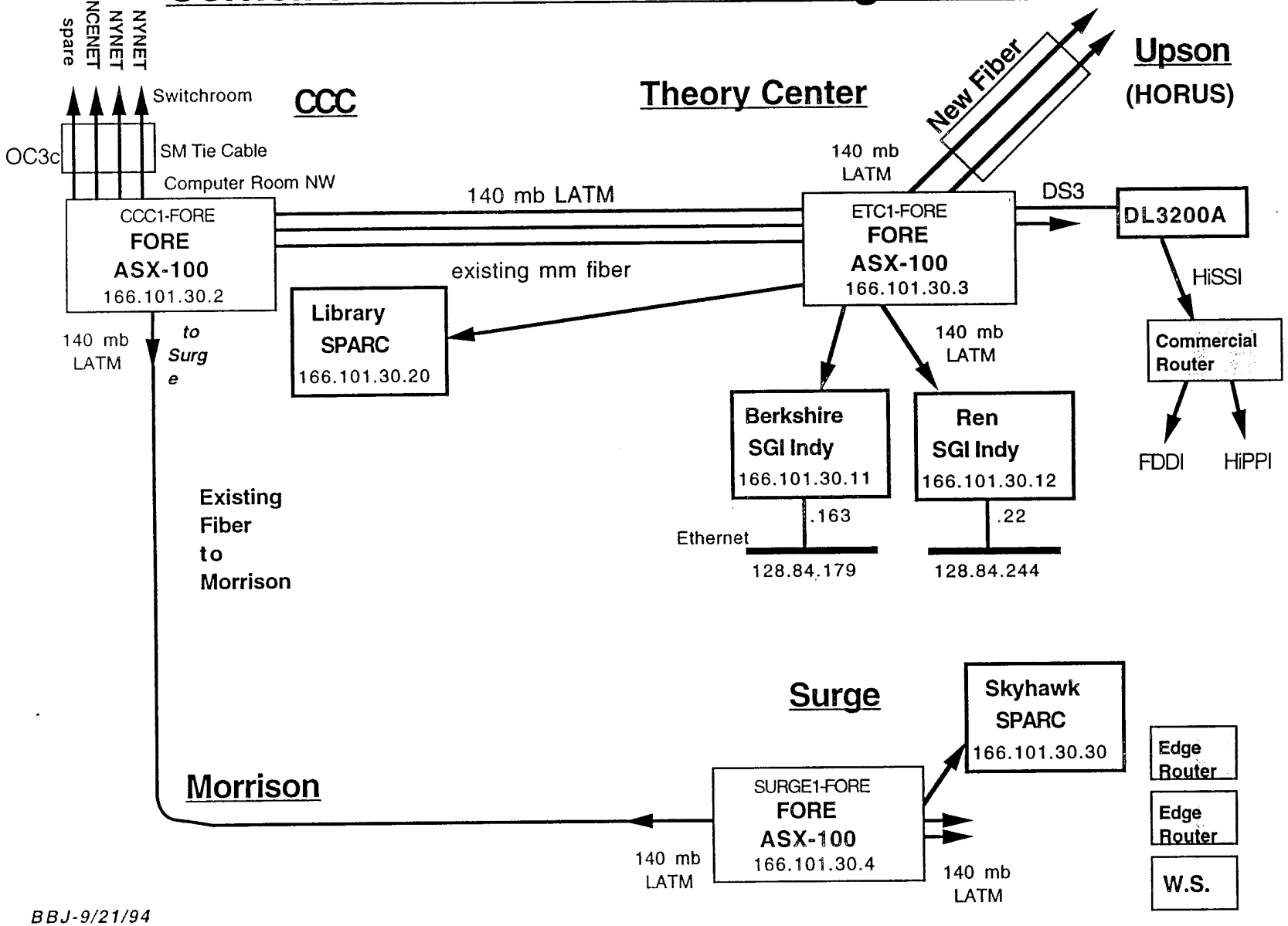
4.1.6.2 Packet video teleconferencing

See section 4.1.3.

4.1.6.3 ATM LAN architecture

Our ATM LAN architecture is described in the following diagram:

Cornell's FORE ATM Switch Configuration



BBJ-9/21/94

4.1.6.4 Network load generation and analysis of loads

Due to networking difficulties these tests were not performed for inclusion in this report. They will be performed at a later date.

4.1.6.5 Proof of concept of digital libraries

See section 4.1.4.

Conclusions

While much has been said about the development of a national, or even global, information infrastructure, the broad focus has been on providing stable, ubiquitous connectivity with an assumption that, while aggregate bandwidth requirements are going to be large, individual requirements will be relatively modest. Within the context of NYNET, both Cornell and Rome Laboratory have recognized the importance of challenging these assumptions as insufficient. Cornell has a unique range of applications, from high-speed point-to-point "virtual reality" or telecollaboration environments to massive aggregate data servers to commodity phone and video services.

Our goal has been to create a testbed which assumes a world where there is a large range of networking service requirements, varying in terms of individual and aggregate bandwidth, reliability requirements and accounting. In this initial project, we have identified many issues, from monitoring and managing congestion and packet loss, to interoperability issues among ATM sites and between ATM and other protocols. Much more needs to be done before the community will be able to build effective tools for network management. While outside the scope of this study, we note that more and more, network management systems must become 'aware' of the requirements of the applications which run across them; this is especially true for command and control applications, as well as other applications which must run in "real time".

While recognizing the importance of fundamental research on network traffic models, packet sizes and protocols, we have taken a very pragmatic approach, especially given our plans (already underway) to move as quickly as we can to an integrated ATM-infrastructure for voice, computer data, and video. With the Theory Center's strong emphasis on absolute performance for high performance computing applications, we have been able to push in areas which will filter "down" to the typical user over the next few years; at the same time, Cornell Information Technologies, with its leadership in video and collaboration technologies, has demonstrated that some of the most demanding applications will come from the use of high speed networks in telecollaboration, distance learning and other mixes of video, computer display technology and people. Thus, while exploiting the availability of HPC applications in our ATM studies, we are certain that our findings are consistent with a far broader range of applications and user communities.

Appendix A: Cells in Frames (CIF)

From a maillist posting July, 1994

R. Cogger, Cornell University

Why not run ATM cells in Ethernet frames?

Background: At Cornell, we have over 8000 ports of 10BASE-T ethernet installed, going to 10,000 by the end of another year. We want to get ATM to all of those desktops as soon as possible.

Solutions being offered or proposed by various vendors promise 25Mbps or 51Mbps or more, but the most optimistic pricing I've heard, projecting for a year or two in the future is about $3-400/port on a mux plus $2-300 for the workstation adapter. Well, that means $5-700 per desktop or $5-7,000,000 for our 10,000 desktops. No way will we be able to afford that much. Also, some of those desktop systems have built-in ethernet and no slot for a new ATM adapter.

So, my proposal to prospective vendors is: build a 24-port mux (10BASE-T ports) with a 155Mbps ATM port to connect via fiber to a switch. Price it so that the volume discounted cost for us is $100/port, and we will buy approximately 10,000 ports. I think there are probably many campuses with similar needs.

Here is how it should work:

1. The mux needs to be as simple as possible to make the cost target. So it should do as little as possible.

2. Rule 1. -- Only one node (workstation) on a port, so it is a point to point link. No contention, no collisions, etc.

3. AAL is done in software at the workstation (a new driver); SAR is distributed between the workstation and the mux (see below).

4. The UNI is available via appropriate API to Native-mode ATM applications. IP etc. run over ATM in *exactly* the same way they do with the 25Mbps etc. desktop solutions. The same native ATM applications run, although not as fast, of course with only half-duplex 10Mbps available. Performance will be CPU-limited also for some workstations, but remember it is only 10Mbps.

5. A minimum size ethernet frame is 64 bytes with a 46 byte payload which means that the minimum frame for a cell would be 66 bytes, unless some games were played such as use the type bytes for payload.

6. How about the cell header? Implement the mux as a promiscuous listener. For frames sent toward the mux, put the cell header in the ethernet destination address. Perhaps the mux can calculate the HEC to off load the workstation a little. For frames sent by the mux toward the workstation, the cell header goes in the source address.

7. Rule 2. -- More than one cell payload can go in a frame, but they must all be for the same VC; hence they all have the same cell header; hence the cell header only needs to be supplied once, as above. There is no overhead on the wire for sending data as cells. There is very little compute overhead for the workstation to perform the SAR.

8. For workstation originated data, the workstation can trade off efficiency versus latency by choice of how many cells to put in a frame.

9. For data coming in from the ATM switch and being sent on the ethernet link in frames by the mux, there will (possibly) need to be some signaling at VC setup to set parameters for how the mux frames the cells. The mux, receiving a cell from the high-speed link can dally to see if another for the same VC arrives "soon enough" to be packaged in the same frame. There can be a max number of cells before sending the frame.

10. VC's connecting ports on the same mux could have cells (frames?) switched internally in the mux or they could simply be sent as usual on the high-speed link so the switch at the other end can send them back. Probably it's more important to keep the cost of the mux down by keeping it simple than it is to economize on bandwidth used on the link to the switch.

Other comments:

This proposal is an alternative to ethernet frame switching, which seems costly, and to several other forms of "preserving investment in the installed base." We've spent the last 5 years getting those 10,000 folks hooked up, learning to operate it, manage it, etc. The next step needs to be directly to ATM to the desktop for everyone.

Whether to use such an infrastructure to do LAN emulation, or (my favorite) router emulation for existing applications can be debated in exactly the same terms as for the 25Mbps, 51Mbps, etc. solutions.
What about performance? For someone converting from a shared ethernet, they should see a 10 or 20 fold improvement, depending on many things. They should also benefit greatly from being able to get QoS facilities, especially for low bandwidth services such as voice. It should be possible to transmit middle-resolution MPEG at 6 Mbps, unless the workstation is a lower-end PC. Anyway, the point is to get early, massive deployment, affordably. I assume we would also be using the 25Mbps or 51Mbps solutions for those able to afford more $. In effect, being able to show folks the benefits of ATM to the desktop may help to sell them on the need for higher priced, higher speed ATM.

Appendix B: Briefing Outline

Building a High-Speed Networking Infrastructure

- **Project Status Briefing and Demonstration**
  - Rome Laboratories
  - September 21, 1994
  - Presented By:
    - » Jeffrey A. Silber, Dir. Admin & Op. Sup., CTC
    - » Bruce B. Johnson, Lead Network Engineer, CU
    - » Richard Gillilan, Visualization Specialist, CTC

Building a High-Speed Networking Infrastructure

- **Agenda**
  - Project History and Description
    - » Jeffrey Silber
  - Detailed Status
    - » All Cornell Participants
  - Demonstrations
    - » Richard Gillilan
    - » Bruce Johnson
    - » Jeffrey Silber

Building a High-Speed Networking Infrastructure

- **Project History**
  - Collaboration between Cornell and Rome Labs
    - » Theory Center
      - Description of Center
      - Management of Center
      - Computing and Network Environment
    - » Network Resources
      - Description of CIT
      - NR Activities
    - » Computer Science
      - ISIS/HORUS
      - SP2, High Performance Switch

Building a High-Speed Networking Infrastructure

- **Project History**
  - NYNET
    - » NYNET Partners and Structure
      - Original Members
        - **NYNEX**
        - **Columbia**
        - **Cornell**
        - **Polytechnic University**
        - **Rome Laboratories**
        - **Syracuse University, NPAC**
      - Applications Committee
      - Infrastructure Committee

Building a High-Speed Networking Infrastructure

- NYNET
  - » NYNET Physical Infrastructure
    - Network Wide
    - Local Cornell Infrastructure
  - » NYNET Applications
    - Summary

Building a High-Speed Networking Infrastructure

- **Project History**
  - Response to Rome Labs BAA 90-04
    - » First Step in a new Cornell–Rome Labs Relationship
    - » One Portion of a Far Broader Scope of Activities
    - » Purpose of this Briefing
      - Final Report

Building a High-Speed Networking Infrastructure

- **Project Status and Findings**
- **TCP/IP, ATM Issues**
  - TCP/IP–ATM Interaction
  - ATM Technology at Cornell
  - ATM Intraoperability and Lan Interconnection
  - Applications
    - » Virtual Reality and Real Time Imaging
    - » Information Servers: Mosaic, Digital Libraries

Building a High-Speed Networking Infrastructure

- **Project Status and Findings**
- **NYNET Network Management Infrastructure**
  - Fore Switches
  - Netview/6000
  - Future Issues
- **Multiparty Conferencing over Packet Video**
  - Technical Details
  - User Interface
  - Reflector
  - Slide Window

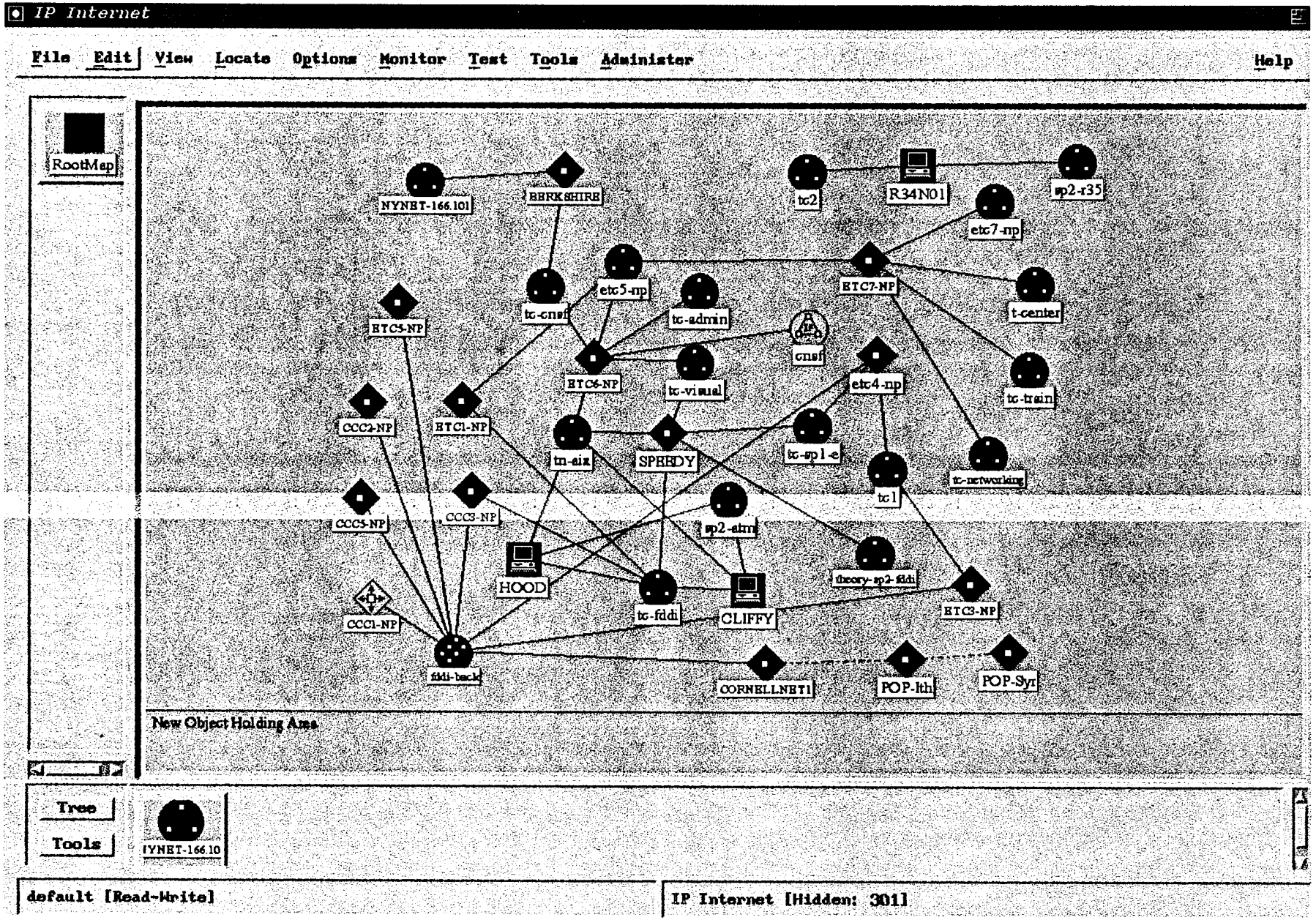
Building a High-Speed Networking Infrastructure

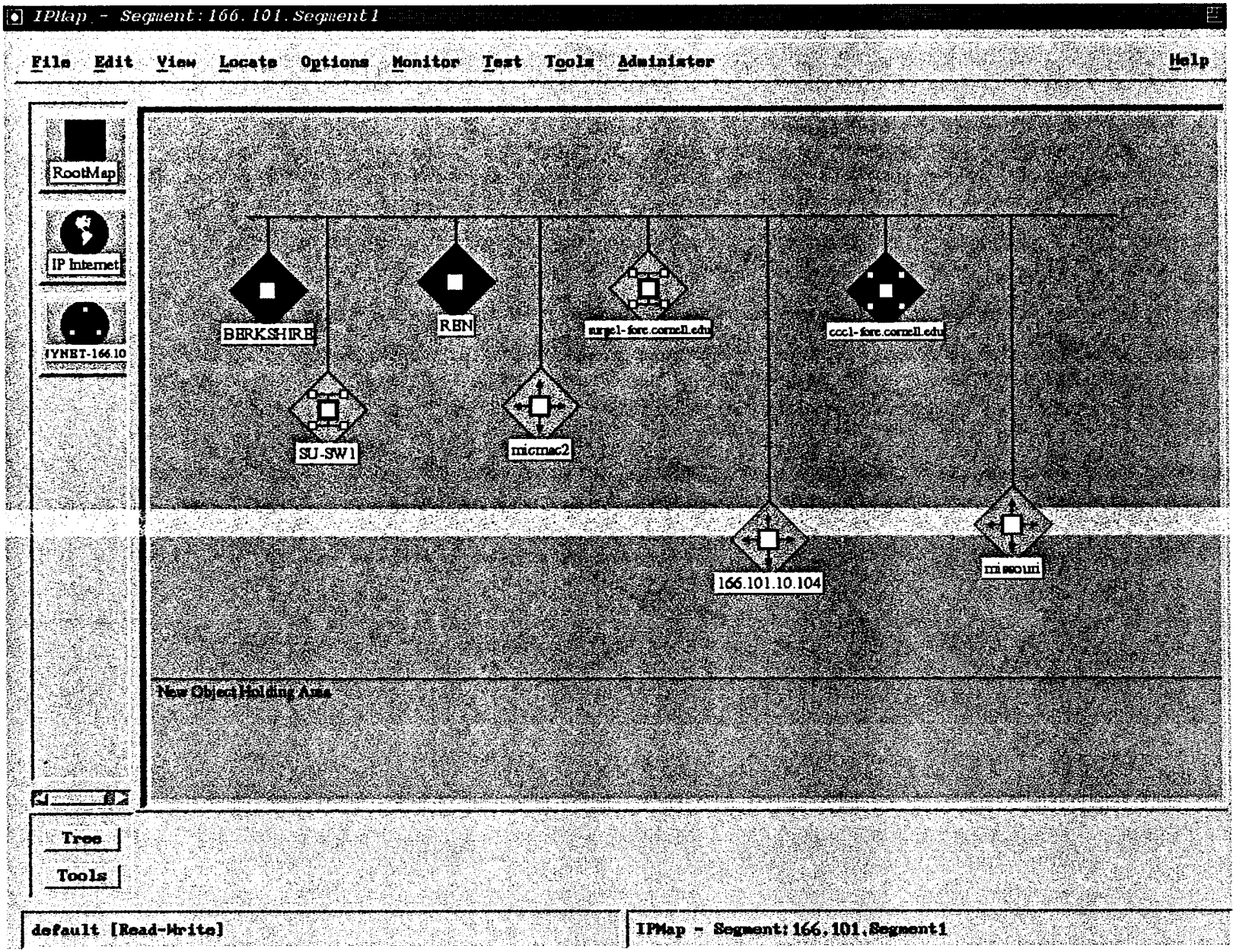
- **Project Status and Findings**
- **Medical Imaging Application**
  - Capabilities
  - Image Transport (via WWW, CU-SeeME)
  - Virtual Reality and Real Time Imaging

Building a High-Speed Networking Infrastructure

- **Demonstrations**
  - CU-SeeMe
  - Virtual Reality Application
  - Information Server (WWW)
  - Digital Libraries

Appendix C: Network Monitor


IP Internet
File
Edit
View
Locate
Options
Monitor
Test
Tools
Administer
Help
RootMap
NYNET-166.101
BERKSHIRE
tc2
R34N01
sp2-r35
etc7-rp
ETC7-NP
t-center
etc5-rp
tc-gnsf
tc-admin
ETC5-NP
gnsf
ETC6-NP
tc-visual
etc4-rp
tc-train
CCC2-NP
ETC1-NP
tn-elx
SPEEDY
tc-sp1-e
tc-networking
tc1
CCC5-NP
CCC3-NP
sp2-atm
HOOD
theory-sp2-fddi
CCC1-NP
tc-fddi
CLIFFY
ETC3-NP
fddi-back
CORNELLNET1
POP-Ith
POP-Syr
New Object Holding Area
Tree
Tools
NYNET-166.10
default [Read-Write]
IP Internet [Hidden: 301]


IPMap - Segment:166.101.Segment1
File Edit View Locate Options Monitor Test Tools Administer Help
RootMap
IP Internet
CYNET-166.10
BERKSHIRE
RBN
surge1-fore.cornell.edu
ccc1-fore.cornell.edu
SU-SW1
micmac2
166.101.10.104
missouri
New Object Holding Area
Tree
Tools
default [Read-Write]
IPMap - Segment:166.101.Segment1

Rome Laboratory

Customer Satisfaction Survey

RL-TR-_____

Please complete this survey, and mail to RL/IMPS, 26 Electronic Pky, Griffiss AFB NY 13441-4514. Your assessment and feedback regarding this technical report will allow Rome Laboratory to have a vehicle to continuously improve our methods of research, publication, and customer satisfaction. Your assistance is greatly appreciated.
Thank You

_____

_____

Organization Name:_____(Optional)

Organization POC: _____(Optional)

Address:_____

1. On a scale of 1 to 5 how would you rate the technology developed under this research?

5-Extremely Useful 1-Not Useful/Wasteful

Rating_____

Please use the space below to comment on your rating. Please suggest improvements. Use the back of this sheet if necessary.

2. Do any specific areas of the report stand out as exceptional?

Yes___ No_____

If yes, please identify the area(s), and comment on what aspects make them "stand out."

3. Do any specific areas of the report stand out as inferior?

Yes___ No___

If yes, please identify the area(s), and comment on what aspects make them "stand out."

4. Please utilize the space below to comment on any other aspects of the report. Comments on both technical content and reporting format are desired.

MISSION

OF

ROME LABORATORY

Mission. The mission of Rome Laboratory is to advance the science and technologies of command, control, communications and intelligence and to transition them into systems to meet customer needs. To achieve this, Rome Lab:

a. Conducts vigorous research, development and test programs in all applicable technologies;

b. Transitions technology to current and future systems to improve operational capability, readiness, and supportability;

c. Provides a full range of technical support to Air Force Materiel Command product centers and other Air Force organizations;

d. Promotes transfer of technology to the private sector;

e. Maintains leading edge technological expertise in the areas of surveillance, communications, command and control, intelligence, reliability science, electro-magnetic technology, photonics, signal processing, and computational science.

The thrust areas of technical competence include: Surveillance, Communications, Command and Control, Intelligence, Signal Processing, Computer Science and Technology, Electromagnetic Technology, Photonics and Reliability Sciences.